

Improving Robustness in Grammars

Dan Flickinger

DELPH-IN Summit 2023

Galicia, Spain

29 June 2023

Strategies for increasing robustness

- Expand curated lexicon (e.g. 3000 most frequent verbs in BNC)
- Add generic entries triggered by POS tagger or chart mapping
- Add chart mapping rules for dates, measures, numerals, etc.
- Add grammar constructions for mild but frequent ungrammaticality
- Overcome computing resource limits

Shrink feature structures via addition of *min* types

Prune parse chart via supertagging or PCFGs

Make more RAM available for longer sentences

Generic entries triggered by POS tagger

```
generic_trans_verb_past := v_np*_pa-unk_le &  
  [ ORTH < "_generic_vbd_" >,  
    TOKENS.+LIST < [ +TNT.+TAGS.FIRST "VBD" ] > ].
```

```
generic_adj_compar := aj_-_i-cmp-unk_le &  
  [ ORTH < "_generic_jjr_" >,  
    TOKENS.+LIST < [ +TNT.+TAGS.FIRST "JJR" ] > ].
```

Generic entry triggered by chart mapping

Chart mapping rule for '94:

```
apostrophe_year_ne_tmr := basic_ne_tmt &  
[ +INPUT < [ +FORM ^['][0-9][0-9](+|-|)?$ ] > ,  
  +OUTPUT < [ +CLASS year_ne ] > ].
```

Generic lexical entry:

```
generic_year_ne := n_-_pn-yoc-gen_le &  
[ ORTH < "_generic_year_ne_" > ,  
  TOKENS.+LIST < [ +CLASS year_ne ] > ].
```

Constructions for mild ungrammaticality

- Bare singular count noun, in recipes or titles

Place book on cart

- Slashed VP in dictionary definitions

rely: to depend on